**Coursework for M4MBA3 Optimisation**

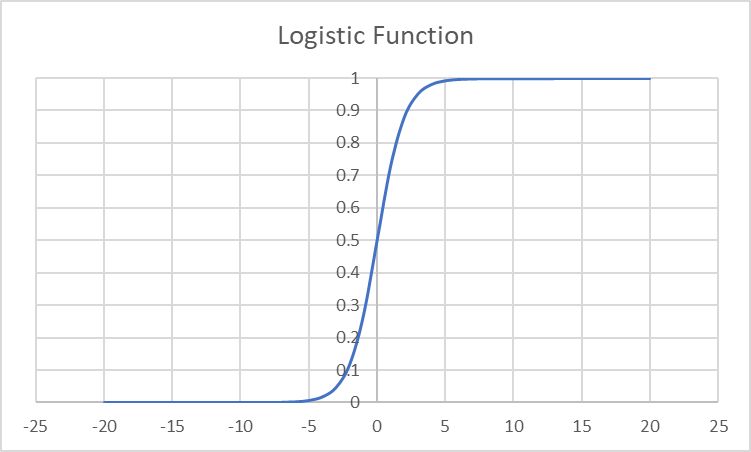
**Using Logistic Regression Using Excel**

**1. Background:**

Logistic Regression (LR) – also known as Logit Regression or Logit Model - is a statistical estimation method. Given a set of data, which classifies each data point in a binary manner, the aim is to find an exponential that can learn the patterns and return probabilities about future patterns.

As an example, assume that you are given a set of data about students that consists of if they have passed the MN50642 Optimisation unit (, with 0 meaning fail and 1 meaning pass), and how much time they have spent reading the textbook (), solving examples on paper (), and solving problems on a computer ().

LR makes use of the Logistic Function , a well-behaved function with useful mathematical properties, plotted in the figure below. Simply put, the Logistic Function is a continuous approximation of a step-function, better known as an if function. It has widespread uses in Machine Learning.



The Logistic Function can be used as an estimator of the probability of an event since it returns values between 0 and 1. However, it is based only on one input, . We would like to find a linear function to replace the in the Logistic Function, of the form . The modified Logistic Function then can be used to estimate the probability that student will pass. Consequently, is the estimated probability that student will fail.

The problem boils down to finding the optimal values for that will return accurate predictions. The corresponding optimisation model is:

(LR-P1)

(1)

(2)

LR-P1 “maximum likelihood” model, which aims to maximise the probability that the observed pass / fail values were the most likely outcome. Note that LR-P1 is never infeasible, and the objective function is concave, making it easy to implement and solve.

However, the objective function can cause problems for large data sets. Multiplying many numbers that are in the range (0,1) results in an objective function that is very small, which can cause numerical problems. An equivalent approach is to use a “log-transform”, that is, using a logarithm function on the objective function, transforming it to a sum. The resulting model is then:

(LR-P2)

(1)

(2)

Note that the logarithm function can be of any base, but it is usually assumed to be base 2, , or 10. The two models are mathematically equivalent, but one may be preferable to the other based on the data on hand.

**2. The Problem:**

You are working for a Mobile Network Provider called Hi5, which aims to provide high speed 5G connection to its users. A brief analysis of the data for the last three months revealed that there is a significant amount of “churn”, users that switch over to other mobile networks. There is no established method of predicting who may churn.

Your manager (Ms Maeby Wright) decides to use her secret weapon, the business analyst she has recently recruited. She provides you with usage data of 100 customers for the last three months, 30 of which have churned. The data consists of the hours of call and GB of data used by each customer in each month. She tells you to find a method of predicting the probability a customer will churn using the given set of data. You decide to use LR, and she agrees with you.

**3. Key Questions from Ms Wright**

1. Implement LR-P1 and LR-P2. Discuss the relative performances of these two models and pick one to use.
2. Comment on the quality of data you are given. What would ideal data look like?
3. Would your model still work if the data consisted of 1,000 customers? How about 10,000 customers?
4. Build a worksheet that will take the optimal solution of the LR to predict the probability of churn for a customer, given the customer’s usage data for the last three months.

**4. Report Specifications**

Ms Wright expects your final report to have the following structure:

1. **Main cover page:** your name and the word count.
2. **2000 words +10%** to address the evaluation criteria given below.
3. **Appendices (at the end)**: you may want to include any relevant content (e.g. details that should not go into the main body).

Sticking to the strict word limit is difficult, but it is an important skill to acquire. Make sure that you write in a concise and focused manner. It should be typed font size 12 and single line spacing. You must include a word count at the start of the report. You report should be aimed at a non-technical audience that may or may not have a general understanding of the field of optimisation.

**You must also provide an Excel file that shows the details of your work.**

**Deadline:**

**Assessment criteria:**

|  |  |  |
| --- | --- | --- |
| **Good report** | **Poor report** | **Mark %** |
| A suitable model has been developed and relevant analytical techniques have been demonstrated to explain the data. | The Excel file with the model is missing. A not suitable model has been chosen and suitable analytical techniques have not been applied for interpreting the data provided | **30%** |
| A thorough listing and discussion of the assumptions and limitations of your model. | There is no discussion on the assumptions and limitations of the model. The limitations and/or assumptions are not valid. | **20%** |
| Illustrating clear and deep understanding of how to use an analytical model in business decision making related to the key questions. | The answers to the questions are not clear and/or comprehensive, and do not illustrate deep understanding of the actual questions. There are some answers missing and/or are wrong. | **30%** |
| Format, quality of expression, clarity of ideas. | The format is not clear and the structure of the document is not easy to understand and follow. | **20%** |

**DATA:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Month 1** | | |  | | **Month 2** | | | **Month 3** | |
| **Customer** | **Churn** | | **Call (hours)** | **Data (GB)** | | **Call (hours)** | | **Data (GB)** | **Call (hours)** | | **Data (GB)** | |
| 1 | Yes | | 0.67 | 0.85 | | 1.55 | | 2.28 | 2.35 | | 1.85 | |
| 2 | Yes | | 0.86 | 0.67 | | 1.03 | | 1.52 | 1.38 | | 2.79 | |
| 3 | Yes | | 0.94 | 1.27 | | 2.31 | | 1.98 | 3.33 | | 1.98 | |
| 4 | Yes | | 1.10 | 1.48 | | 2.30 | | 1.31 | 3.37 | | 2.86 | |
| 5 | Yes | | 0.94 | 0.98 | | 1.53 | | 1.03 | 1.23 | | 1.60 | |
| 6 | Yes | | 0.44 | 1.35 | | 1.34 | | 2.51 | 3.18 | | 3.43 | |
| 7 | Yes | | 0.17 | 1.28 | | 2.44 | | 2.22 | 1.16 | | 3.27 | |
| 8 | Yes | | 1.04 | 1.49 | | 2.30 | | 2.85 | 1.02 | | 1.53 | |
| 9 | Yes | | 0.95 | 0.54 | | 1.81 | | 2.59 | 3.41 | | 1.88 | |
| 10 | Yes | | 0.66 | 0.64 | | 2.27 | | 2.54 | 1.67 | | 2.72 | |
| 11 | Yes | | 0.54 | 0.60 | | 2.02 | | 1.70 | 1.47 | | 2.61 | |
| 12 | Yes | | 0.19 | 0.86 | | 0.89 | | 1.82 | 2.23 | | 2.76 | |
| 13 | Yes | | 0.86 | 1.05 | | 1.85 | | 1.57 | 1.23 | | 1.85 | |
| 14 | Yes | | 0.52 | 1.06 | | 2.05 | | 2.11 | 3.47 | | 2.24 | |
| 15 | Yes | | 0.30 | 1.42 | | 1.35 | | 1.19 | 1.13 | | 2.16 | |
| 16 | Yes | | 0.78 | 0.95 | | 1.59 | | 1.54 | 3.42 | | 2.25 | |
| 17 | Yes | | 0.65 | 1.41 | | 1.56 | | 1.84 | 1.65 | | 2.32 | |
| 18 | Yes | | 0.67 | 0.53 | | 1.31 | | 2.98 | 3.34 | | 2.12 | |
| 19 | Yes | | 1.10 | 0.56 | | 1.83 | | 2.76 | 3.42 | | 2.99 | |
| 20 | Yes | | 0.69 | 1.44 | | 1.03 | | 2.68 | 1.43 | | 2.18 | |
| 21 | Yes | | 0.98 | 1.33 | | 0.72 | | 1.26 | 2.52 | | 3.21 | |
| 22 | Yes | | 0.51 | 1.23 | | 1.69 | | 1.28 | 2.99 | | 2.11 | |
| 23 | Yes | | 1.09 | 1.39 | | 0.53 | | 2.08 | 1.59 | | 2.03 | |
| 24 | Yes | | 0.94 | 1.45 | | 1.48 | | 2.44 | 1.54 | | 2.02 | |
| 25 | Yes | | 0.48 | 1.43 | | 1.38 | | 1.26 | 3.30 | | 3.21 | |
| 26 | Yes | | 0.12 | 0.60 | | 1.03 | | 1.43 | 2.24 | | 2.41 | |
| 27 | Yes | | 0.58 | 1.10 | | 1.14 | | 2.55 | 2.38 | | 2.51 | |
| 28 | Yes | | 0.17 | 0.63 | | 1.88 | | 1.81 | 2.34 | | 2.45 | |
| 29 | Yes | | 0.98 | 0.80 | | 0.85 | | 1.55 | 1.04 | | 1.68 | |
| 30 | Yes | | 0.28 | 1.11 | | 1.65 | | 2.48 | 2.53 | | 2.99 | |
| 31 | No | | 1.24 | 2.21 | | 0.72 | | 2.72 | 2.31 | | 1.51 | |
| 32 | No | | 2.49 | 2.75 | | 2.13 | | 1.99 | 0.73 | | 1.87 | |
| 33 | No | | 0.92 | 1.77 | | 1.53 | | 2.98 | 2.37 | | 1.60 | |
| 34 | No | | 0.70 | 1.43 | | 1.32 | | 1.62 | 2.48 | | 2.61 | |
| 35 | No | | 1.02 | 1.95 | | 1.52 | | 2.64 | 1.71 | | 2.60 | |
| 36 | No | | 2.04 | 2.50 | | 1.77 | | 1.27 | 0.83 | | 1.73 | |
| 37 | No | | 2.13 | 2.00 | | 2.11 | | 1.58 | 1.64 | | 2.35 | |
| 38 | No | | 0.51 | 1.98 | | 0.95 | | 1.71 | 1.45 | | 1.46 | |
| 39 | No | | 0.74 | 1.67 | | 2.17 | | 1.74 | 1.16 | | 2.92 | |
| 40 | No | | 0.78 | 1.27 | | 0.64 | | 1.54 | 1.64 | | 2.61 | |
| 41 | No | | 2.00 | 1.30 | | 0.91 | | 2.56 | 0.68 | | 1.45 | |
| 42 | No | | 1.31 | 1.26 | | 0.96 | | 2.50 | 1.42 | | 1.32 | |
| 43 | No | | 0.68 | 1.97 | | 2.29 | | 1.05 | 1.77 | | 2.76 | |
| 44 | No | | 2.11 | 2.51 | | 0.74 | | 1.97 | 1.54 | | 1.20 | |
| 45 | No | | 1.14 | 2.54 | | 0.94 | | 1.02 | 1.63 | | 1.70 | |
| 46 | No | | 1.19 | 2.54 | | 1.38 | | 2.05 | 1.57 | | 1.40 | |
| 47 | No | | 1.56 | 2.33 | | 2.19 | | 1.46 | 1.22 | | 1.90 | |
| 48 | No | | 1.03 | 1.98 | | 0.56 | | 2.29 | 1.52 | | 1.98 | |
| 49 | No | | 1.45 | 1.09 | | 1.68 | | 2.88 | 1.17 | | 2.81 | |
| 50 | No | | 2.01 | 1.09 | | 0.85 | | 1.82 | 2.41 | | 2.38 | |
| 51 | No | | 1.79 | 1.84 | | 1.57 | | 2.65 | 1.71 | | 2.25 | |
| 52 | No | | 2.10 | 1.82 | | 0.94 | | 1.27 | 1.95 | | 2.25 | |
| 53 | No | | 2.27 | 1.24 | | 2.07 | | 1.35 | 0.69 | | 2.66 | |
| 54 | No | | 1.04 | 1.45 | | 1.92 | | 1.50 | 1.49 | | 1.25 | |
| 55 | No | | 1.76 | 1.10 | | 2.37 | | 2.22 | 1.36 | | 1.55 | |
| 56 | No | | 1.89 | 1.49 | | 1.78 | | 1.20 | 1.73 | | 2.78 | |
| 57 | No | | 2.40 | 2.07 | | 1.27 | | 1.98 | 0.82 | | 1.61 | |
| 58 | No | | 1.05 | 2.84 | | 1.20 | | 2.00 | 1.29 | | 1.55 | |
| 59 | No | | 1.43 | 2.95 | | 1.44 | | 2.89 | 1.21 | | 2.76 | |
| 60 | No | | 1.56 | 1.83 | | 2.31 | | 1.66 | 1.37 | | 1.04 | |
| 61 | No | | 0.79 | 1.04 | | 1.91 | | 1.93 | 0.98 | | 1.49 | |
| 62 | No | | 0.53 | 2.28 | | 0.82 | | 1.18 | 1.79 | | 1.08 | |
| 63 | No | | 1.34 | 1.99 | | 1.76 | | 1.41 | 1.82 | | 2.92 | |
| 64 | No | | 2.09 | 2.39 | | 1.46 | | 1.04 | 1.62 | | 1.47 | |
| 65 | No | | 1.80 | 2.78 | | 1.43 | | 1.25 | 0.80 | | 2.75 | |
| 66 | No | | 0.91 | 2.87 | | 2.02 | | 2.78 | 1.28 | | 2.08 | |
| 67 | No | | 0.96 | 2.36 | | 1.62 | | 1.25 | 1.95 | | 1.48 | |
| 68 | No | | 0.81 | 2.87 | | 1.13 | | 2.08 | 2.19 | | 1.63 | |
| 69 | No | | 1.73 | 1.60 | | 2.14 | | 2.40 | 2.15 | | 1.36 | |
| 70 | No | | 0.91 | 1.01 | | 2.18 | | 1.99 | 0.56 | | 1.14 | |
| 71 | No | | 1.74 | 2.99 | | 0.96 | | 1.32 | 0.62 | | 2.42 | |
| 72 | No | | 0.55 | 2.82 | | 2.00 | | 1.47 | 1.24 | | 1.77 | |
| 73 | No | | 1.58 | 1.80 | | 1.65 | | 1.14 | 0.99 | | 2.20 | |
| 74 | No | | 1.97 | 2.03 | | 1.84 | | 1.33 | 2.26 | | 2.35 | |
| 75 | No | | 2.44 | 1.63 | | 2.11 | | 1.72 | 1.59 | | 2.93 | |
| 76 | No | | 1.41 | 2.52 | | 0.71 | | 2.38 | 2.11 | | 2.09 | |
| 77 | No | | 0.83 | 1.75 | | 1.24 | | 2.40 | 2.35 | | 2.92 | |
| 78 | No | | 0.95 | 1.49 | | 2.18 | | 1.37 | 1.24 | | 1.71 | |
| 79 | No | | 1.58 | 2.73 | | 2.23 | | 1.31 | 1.96 | | 2.73 | |
| 80 | No | | 1.14 | 2.45 | | 1.59 | | 2.47 | 0.55 | | 1.37 | |
| 81 | No | | 0.87 | 1.28 | | 1.44 | | 1.75 | 2.11 | | 2.21 | |
| 82 | No | | 1.19 | 1.75 | | 0.73 | | 2.77 | 1.65 | | 1.13 | |
| 83 | No | | 1.68 | 1.60 | | 2.45 | | 2.86 | 0.95 | | 2.64 | |
| 84 | No | | 1.90 | 1.85 | | 1.40 | | 1.33 | 2.39 | | 2.82 | |
| 85 | No | | 0.51 | 2.74 | | 0.58 | | 2.22 | 1.75 | | 3.00 | |
| 86 | No | | 1.01 | 2.17 | | 2.03 | | 1.98 | 2.03 | | 1.92 | |
| 87 | No | | 0.83 | 1.15 | | 1.87 | | 2.51 | 1.11 | | 1.08 | |
| 88 | No | | 1.63 | 2.79 | | 1.38 | | 1.48 | 2.05 | | 1.46 | |
| 89 | No | | 0.83 | 2.50 | | 1.39 | | 2.20 | 1.03 | | 2.24 | |
| 90 | No | | 1.64 | 2.09 | | 2.28 | | 1.77 | 2.12 | | 2.01 | |
| 91 | No | | 1.42 | 2.97 | | 1.34 | | 1.11 | 2.12 | | 1.17 | |
| 92 | No | | 1.29 | 1.05 | | 1.79 | | 1.54 | 1.79 | | 2.65 | |
| 93 | No | | 0.77 | 2.94 | | 0.95 | | 2.08 | 1.02 | | 1.30 | |
| 94 | No | | 1.12 | 2.94 | | 1.08 | | 2.23 | 2.42 | | 1.45 | |
| 95 | No | | 0.69 | 1.88 | | 1.18 | | 2.40 | 1.76 | | 1.24 | |
| 96 | No | | 1.39 | 2.33 | | 2.22 | | 1.94 | 1.56 | | 2.93 | |
| 97 | No | | 1.13 | 2.93 | | 1.85 | | 2.59 | 0.52 | | 2.79 | |
| 98 | No | | 1.78 | 2.50 | | 1.21 | | 1.57 | 0.50 | | 2.23 | |
| 99 | No | | 2.39 | 1.71 | | 1.21 | | 1.77 | 2.47 | | 2.91 | |
| 100 | No | | 1.45 | 2.56 | | 2.40 | | 2.76 | 1.01 | | 1.35 | |